2016

Adding Quantitative Corpus-Driven Analysis to Qualitative Discourse Analysis: Determining the Aboutness of Writing Center Talk

Jo Mackiewicz
Iowa State University, jomack@iastate.edu

Isabelle Thompson
Auburn University

Adding Quantitative Corpus-Driven Analysis to Qualitative Discourse Analysis: Determining the Aboutness of Writing Center Talk

Abstract
We discuss the benefits of using corpus linguistic analysis, a quantita- tive method for determining the "aboutness" of talk, in conjunction with discourse analysis in order to understand writing center talk at a micro- and macrolevel. We exemplify this mixed-method approach by examining a specialized corpus of 20 writing center conferences totaling more than 75,000 words. Our analysis also uncovered words that differentiated writing center talk from reference corpora and thus helped reveal the aboutness of the writing center talk. For example, student writers said "I don't know" far more frequently than any other 4-gram, and tutors said "You're going to" far more frequently than other 4-grams. We close by discussing the possibility of creating a corpus of writing center talk that researchers could use to ask and answer a broad range of research question

Disciplines
Educational Assessment, Evaluation, and Research | Rhetoric and Composition

Comments
This article is published as “Adding Quantitative Corpus-Driven Analysis to Qualitative Discourse Analysis: Determining the Aboutness of Writing Center Talk.” The Writing Center Journal 35 (2016): 187–225. (First author, with Isabelle Thompson, 75%). Posted with permission.

Jo Mackiewicz & Isabelle Thompson

Adding Quantitative Corpus-Driven Analysis to Qualitative Discourse Analysis: Determining the Aboutness of Writing Center Talk

Abstract

We discuss the benefits of using corpus linguistic analysis, a quantitative method for determining the "aboutness" of talk, in conjunction with discourse analysis in order to understand writing center talk at a micro- and macrolevel. We exemplify this mixed-method approach by examining a specialized corpus of 20 writing center conferences totaling more than 75,000 words. Our analysis also uncovered words that differentiated writing center talk from reference corpora and thus helped reveal the aboutness of the writing center talk. For example, student writers said "I don't know" far more frequently than any other 4-gram, and tutors said "You're going to" far more frequently than other 4-grams. We close by discussing the possibility of creating a corpus of writing center talk that researchers could use to ask and answer a broad range of research questions.

Since Michael Pemberton's call for research about writing center talk in 2010, *The Writing Center Journal* has published a number of articles reporting analyses of tutors' and student writers' verbal interactions while conferencing (see, for example, Brown, 2010; Corbett, 2011; Severino & Deifell, 2011; Thompson & Mackiewicz, 2014). These studies typically employ some form of qualitative discourse analysis, and a few add quantitative frequency counts. All of these studies provide valuable insights into how tutors attempt to lead and direct student writers' learning and how student writers respond to those attempts. In this article, we discuss the usefulness of another method—corpus linguistic analysis—that provides a different way of examining writing center talk. Corpus linguistic analysis allows quantitative, microlevel examination of words and sequences of words commonly occurring together in a large group of spoken or written texts from the same linguistic register (for example, writing center conferences, classroom discourse, scientific journal articles, or corporations' annual reports). It uncovers words and other linguistic features used so commonly that researchers may otherwise overlook them, and it reveals the keyness of those features in differentiating one register, such as writing center talk, from another. That is, corpus linguistic analysis reveals the "aboutness" of a set of texts (Phillips, 1989). In this article, we discuss the benefit of using corpus linguistic analysis in combination with discourse analysis for a mixed-method approach that pairs a microlevel view with a contextual, macrolevel one for a rich, holistic understanding of language.

Because corpus linguistic analysis is the unfamiliar component of the mixed-method approach that we advocate, we begin by discussing the uses and explaining the benefits of corpus linguistic analysis. Then, we exemplify its application on a 20-conference data set, or corpus, of writing center talk. A corpus is a collection of spoken or written texts collected as a representative sample of a particular linguistic register, or situated language in use. Our more than 75,000-word collection of writing center conferences is a specialized corpus, representative of language use at our particular tutoring site. The conferences were compiled as a sample with the purpose of representing the range of linguistic variation in our writing center. Using Laurence Anthony's (2014) AntConc 3.4.3 concordance software,[1] we measured the following to understand the aboutness of our corpus and to obtain a microlevel view that would sharpen the focus of our macrolevel, discourse analysis:

1 AntConc, developed by Laurence Anthony of Waseda University, is free to download at http://www.laurenceanthony.net. Anthony has supplied substantial support materials for AntConc and the other applications he has developed.

- Tutors' and student writers' conference participation as gauged through word (i.e., token) count
- Tutors' and student writers' range of vocabulary, or type/token ratio (i.e., the ratio of unique word types to tokens)
- Tutors' and student writers' most frequently occurring words, particularly writing-related frequently occurring words
- Tutors' and student writers' key words (words occurring with statistically significant frequency), particularly writing-related key words
- Words collocating with (occurring frequently in proximity to) tutors' and student writers' writing-related key words
- Tutors' and student writers' most frequently occurring four-word lexical bundles (i.e., 4-grams).

As mentioned above, corpus linguistic analysis reveals the aboutness—or the subject matter—of a corpus such as our specialized corpus of writing center conferences. A concept important in a range of disciplines, including information science, psychology, and philosophy, aboutness in linguistics refers to the relation between patterns of language and their content. As Martin Phillips (1989) says, an analysis of "large-scale regularities" in language can reveal how text structure generates the listener's or reader's "psychological perception of subject matter" (p. 7). In other words, analysis at the microlevel—at the level of words and phrases—can help reveal what a text or a corpus is about.

Through corpus linguistic analysis, writing center researchers can begin to ask and answer new research questions. For example, we don't know the extent to which writing center talk differs from everyday conversations or classroom discourse. To what extent do outcomes from writing center conferences arise from such similarities and differences? Also, we don't know the extent to which writing center talk revolves around topics such as developing a main point, organizing and reorganizing content, revising sentences, editing punctuation, and so on. To what extent are tutors and student writers talking about these different components of the writing process? These examples of gaps in the writing center research are just two of a long list of questions that remain unanswered. Our aim here is not to answer these specific questions; instead, it is to show how corpus linguistic analysis can help writing center researchers get the answers to questions such as these. And as we argue here, when combined with discourse analysis, it provides a holistic view of data that is especially informative.

Corpus linguistic analysis provides benefits as a supplement to discourse analysis. First, unlike the process of coding language data by hand, it is not subject to human error. Therefore, it doesn't require

trained coders to ensure replicability. Second, it is objective. It goes beyond human intuition to identify linguistic features worth close examination (McEnery & Wilson, 2001; Hunston, 2002). Third, it can yield generalizable results, as long as the corpus is indeed a representative sample of the larger register.

Two types of corpus linguistics analysis are corpus based and corpus driven (Biber, Conrad, & Reppen, 1998; Butler, 2004; Biber, 2009; Hardie & McEnery, 2010). Researchers use corpus-based analysis, like most quantitative methods, to test hypotheses empirically. Corpus-based analysis, therefore, is deductive. Before they begin their analysis, researchers identify linguistic features relating to their hypotheses. Then, they employ corpus-analysis software to search for and count those features. For example, using theoretical descriptions and possibly previously collected data, researchers can develop a list of linguistic features relating to politeness theory and then determine when and how these features are used and their frequency of occurrence in a corpus of spoken interactions among shoppers and store clerks or among students and classroom teachers. The results would then support or deny the hypotheses derived from the linguistic features associated with politeness theory.

In contrast, corpus-driven analysis is inductive, with no predetermined words or collocations of words as search terms. Rather than driving the research, key words and collocations are derived from the corpus through a variety of quantitative analyses. Theory, then, does not exist before the corpus-driven analysis but is developed from it. In some ways, corpus-driven analysis performs the typically qualitative function of searching out what appears important in a collection of texts. In our corpus-driven analysis, the quantitative analysis identifies the salient features to be examined by a subsequent qualitative discourse analysis.

In the rest of this section, we discuss studies that employ corpus analysis to examine spoken language and thus illuminate our study of writing center conferencing. Focusing on topics also investigated by writing center researchers, these studies reveal the benefits of corpus-driven analysis or corpus-based analysis typically combined with discourse analysis in order to understand writing center talk at both a micro- and macrolevel.

Examining the effectiveness of problem-based learning for teaching in veterinary medicine, while simultaneously investigating the usefulness of a corpus-driven approach to veterinary pedagogy research, Ana DaSilva & Reg Dennick (2010) used a corpus-driven approach exclusively (with no follow up discourse analysis). They analyzed veterinary medicine students' use of appropriate medical vocabulary and

verbal indications of reasoning, explaining, and questioning across three sessions of problem-based learning focused on the same case. They determined word frequencies in each session and made cross comparisons. Results indicated not only increases in students' use of medical vocabulary moving toward more exact terms but also an increase in words such as "because," "why," "how," and, "where," and in questions—all of which are verbal expressions of critical thinking. Although DaSilva & Dennick focused entirely on the microlevel, their study indicates how a corpus-driven approach to analyzing spoken language from educational settings can help illuminate the development of students' knowledge and skills.

Other researchers have taken a two-pronged approach to analyzing spoken language, combining corpus analysis with discourse or conversation analysis.[2] Steve Walsh, Tom Morton, & Anne O'Keeffe (2011) analyzed a subcorpus of small-group tutorials from the Limerick–Belfast Corpus of Academic Spoken English. They began with a corpus-driven analysis to "scope out and quantify recurring linguistic features": 1. most frequent words; 2. frequently occurring collocations (i.e., co-occurrences) of words; and 3. key words (p. 326). By comparing the frequently occurring and key words in their subcorpus to the same types of words from the larger corpus, they identified what was "quantitatively distinctive" (p. 328) about the language of small-group tutorials.

They followed up with qualitative analysis—with conversation analysis—to investigate the quantitative findings more closely, and in this way they were able to describe four "speech-exchange systems" (p. 333) beyond the level of the conversational turn: procedural talk, didactic talk, emphatic talk, and argumentative talk. Their combined approach allowed them to see how these speech-exchange systems were "robust" in the subcorpus, operating throughout the small-group tutorials.

In a third study, Jonathon Reinhardt (2010) investigated the use of what he called "directive constructions" (p. 97) in office-hour one-to-one conferences conducted by faculty members and role-played office hours conducted by international graduate teaching assistants. Referring to the faculty members as "experts" and the graduate students as "learners" in terms of teaching as well as disciplinary knowledge, he compared directive constructions in office-hour conferences from two corpora: ITAcorp (learner data) and the Michigan Corpus of Academic Spoken

2 For our purposes, discourse and conversation analysis represent qualitative complements to corpus analysis. For a good description of the similarities and differences between the two, see Wooffitt (2005).

English (expert data). Reinhardt followed up his quantitative analysis with a qualitative analysis of surveys and interviews from three learners. His corpus analysis revealed that the faculty members used a wider range of directive constructions in their office hours. It also showed that the international teaching assistants made less use of inclusive language and tended to use directive constructions that may have restricted students' choices. His qualitative data revealed some potential reasons for the quantitative findings, reasons stemming from context and culture such as experience, length of time in the United States, and gender. But his interviews with the international teaching assistants also made clear that they "wanted to interact with students in a less authoritarian, more egalitarian style in the future" (p. 104). Reinhart's study reveals the benefits of a quantitative corpus-based analysis used in combination with qualitative data collection and analyses.

Finally, Fiona Farr (2003, 2007) analyzed a 80,000-word subcorpus of the Limerick Corpus of Irish English called POTTI (post-observation tutor–trainee interactions) to determine the aboutness of POTTI's spoken-language data: one-to-one conferences between tutors (experienced classroom teachers) and student teachers (graduate students being trained as teachers) conducted after the tutors had observed the students teaching classes. Farr (2007) compared the 50 most frequently occurring words in POTTI to the 50 most frequently occurring words in three other corpora, consisting of casual and academic conversations. She found that the most frequent words in POTTI were similar to the most frequent words in the reference corpora. All of the 10 most frequent words in POTTI also appeared on the list of frequently occurring words in the three comparison corpora. Word frequency, then, did not indicate POTTI's aboutness. However, it is important to note that this finding is not surprising. The most frequently occurring words across spoken and written English registers are similar: function words such as the articles "the" and "a," forms of "be," conjunctions such as "and," and prepositions such as "to" and "of" (Will, 2012).

Farr also compared tutors' and student teachers' most frequently occurring words, and here she found differences between the specialized language of the tutors and that of the student teachers. The words that occurred most frequently on the tutors' list but that did not appear as frequently on the student teachers' list (for example, "this," "are," "if," "now," "say") revealed "the type of informative and rationalizing narrative engaged in by expert speakers" and included hypothetical statements, evaluations, and hedging (2007, p. 247). In contrast, the words occurring most frequently in the student teachers' speech (for ex-

ample, "mm," "right," "ah") functioned as backchannels and suggested uncertainty.

Farr further compared POTTI with the reference corpora to determine key words. Incorporating qualitative data collection and analysis, Farr used tutors' and student teachers' key words to identify five categories of specialized language in POTTI: teaching-related metadiscourse (for example, "lesson," "class"); cognitive and cathartic words, which relate to the learning and reflection that should occur in the conferences ("mean," "feel," "think"); reference words, which occur similarly in casual conversation (personal pronouns, "yourself," "that," personal names); interactivity words, which indicate participation; and hedging words, which indicate hesitation ("like," "sort," "maybe," "just," "some") (2007, pp. 249–250). Farr's key-word analysis of POTTI clearly illustrates the utility of corpus analysis for understanding the aboutness of a given discourse type.

In an earlier article, Farr (2003) described listeners' responses in the one-to-one conferences comprising the POTTI corpus. Through preliminary qualitative analysis, she identified three types of responses that do not take over the conversational floor. First, backchannels (for example, "uhhuh") simply acknowledge the speaker. These minimal responses signal an intention to continue listening. Other minimal responses, for example, "right," "absolutely," "exactly," in contrast, "comment on the content of the talk" (p. 74). Along with these response types, Farr identified a third, overlaps, where the interlocutors talk at the same time, in comparison to conversational floor–taking interruptions, where one interlocutor claims the floor even though the other has not completed his or her conversational turn.

In her corpus-based analysis, Farr determined the frequencies of minimal responses and then analyzed their functions. Minimal responses, she found, vary widely in their purpose: "A range of functions may be performed discretely or concurrently by these tokens and often without hope for the researcher to disambiguate" (2003, p. 77). She also found that minimal responses depend heavily on the rhetorical situation: "*What* is offered as a minimal acknowledgement often matters less than the fact that something is offered" (2003, p. 77; italics in original). She was able to draw more specific conclusions about responses such as "right" and "exactly." She found that they often appeared in the final feedback step of the well-known teacher-questioning sequence of initiation, response, and feedback.

In summary, researchers have combined both corpus-based and corpus-driven analysis and the quantitative data that it generates with qualitative discourse analysis to study spoken language in a variety of

academic settings. Here, we illustrate this mixed-methods approach by analyzing a specialized corpus of writing center talk. In our conclusion, we describe our ongoing research—a process that includes adding more conferences to our specialized corpus as we get them transcribed. Further, we discuss the issues involved with creating a space where writing center researchers could share transcribed conference data. That is, we discuss the possibility of creating a corpus of writing center talk that writing center researchers could use to ask and answer a broader range of research questions.

A Corpus-Driven Analysis

The 20 writing center conferences in our corpus-driven analysis were collected and transcribed from 2005 to 2008 at a large southeastern university. (Two protocols were approved by our IRB—05-130-ET0507 and 07-167-EP0708.) The conferences were conducted by 17 tutors working with 20 different students, all of whom were enrolled in first-year writing courses or second year world literature courses. (For more information about the writing center and the training of the tutors, see Mackiewicz & Thompson, 2015.)

We separated the tutor and the student-writer discourse from the 20 conferences into discrete files, creating 40 files out of the original 20. Then we "cleaned" the 40 files of extraneous words, such as indications of student writers' or tutors' nonverbal behavior and the names identifying the speakers at the turns. During the cleaning, we also ensured that all terms for the same spoken expression were spelled consistently. For example, "uhhuh" became the consistent spelling for "uh-huh" and "uh huh," and "ok" became the consistent spelling for the backchannel "O. K.," "okay," and "OK," while "okay" became the consistent spelling for the term when it is used as an evaluation ("Your proofreading looks okay to me"). The result was a stripped-down corpus of tutor and student dialogue consisting of 75,101 words.

Using Anthony's AntConc 3.4.3, we compiled tutors' and student writers' most frequently occurring words, key words, words collocating with writing-related key words, and n-grams. As mentioned previously, to determine key words—words critical to analyzing the aboutness of our writing center corpus, we needed a reference corpus or corpora. A reference corpus acts as a point of comparison; words that occur unusually frequently in a specialized corpus (such as our writing center corpus) in comparison to the same words as they occur in (or don't occur in) a reference corpus constitute the key words of the specialized corpus. Wanting to compare writing center discourse against spoken English

as people use it more generally, we used as a reference subsections of two large corpora: Manually Annotated Sub-Corpus (MASC) and The Corpus of Contemporary American English (COCA).[3] A subset of the 15-million word Open American National Corpus, MASC consists of spoken and written American English totaling approximately 500,000 words. Also built from spoken and written American English, COCA consists of approximately 450 million words. However, we used only spoken-text data from these corpora (as opposed to the written-text data)—a total of 469,755 words. With this spoken-language data as a reference, we could examine how the writing center talk in our corpus differs from spoken American English more generally.

In the sections that follow, we examine the aboutness of tutor and student writer talk at a microlevel using corpus-driven text analysis (Phillips, 1989; Gozdz-Roszkowski, 2011). We supplement those findings with the macrolevel view—the view of the broader sociolinguistic context—that discourse analysis provides.

Participation gauged through word frequencies. A corpus-driven approach allowed us to better understand tutors' and student writers' relative contributions to their conferences and to gauge participation beyond word count, the measure typically used to evaluate student writer participation and tutor dominance. We used AntConc to measure the extent to which tutors and student writers contributed to their conferences by tallying the types and tokens in their talk. Each unique word in a corpus constitutes a single type. For example, the word "sentence" is one type. The word "sentence," however, occurred 145 times in tutors' discourse and 27 times in student writers' discourse. Each of these 167 occurrences of the word "sentence" constitutes a token. As in previous analyses of writing center discourse, we determined the percent of tokens (the word count) that tutors and student writers contributed to the overall total to gauge their participation.

In addition, we determined the type/token ratio of tutors' and student writers' talk. A type/token ratio helps in understanding the range of vocabulary (i.e., the lexical variety) that tutors and student writers used and, thus, the aboutness of their talk. By calculating the type/token ratio—dividing the number of types by the number of tokens—we could get a crude sense of the ease or difficulty with which

3 Sample data from MASC and COCA are freely available, as are data from other corpora. MASC is available at http://www.anc.org/data/masc/, and COCA is available at http://corpus.byu.edu/coca/. A comprehensive list of corpora (not all free) is available from the Linguist List at http://linguistlist.org/sp/GetWRListings.cfm?wrtypeid=1.

student writers could understand tutors. The closer the result is to 1, the greater the lexical variety of the vocabulary. More than in our study, such a calculation would be more important in analyzing talk from conferences in which one or both participants is learning the conference language.

Across all of the conferences in our writing center corpus, tutors talked more than student writers. They contributed 70% of the tokens. (See Table 1.) This finding is in keeping with what we found in a previous study; in the teaching stage of writing center conferences, tutors contributed 69.9% of the talk (Mackiewicz & Thompson, 2015). That tutors talk more than student writers is a common finding in writing center research (Thonus 1999b, 2002; Corbett, 2011). In calculating type/token ratio, however, we determined that tutors used the same words over and over. The type/token ratio of their talk was .055, meaning that about 5% of their words were unique. Student writers' talk showed somewhat more lexical variety. The type/token ratio of their talk was .084; about 8% of their words were unique. However, as Keith Stuart (2006) and Norbert Schmitt (2000) pointed out and as Table 1 shows, student writers contributed fewer words to the conferences, (22,512), increasing the likelihood of a larger type/token ratio. The type/token ratio of tutors' talk indicates that they used a fairly narrow range of vocabulary, a narrowness of range that could help many native speakers as well non-native speakers comprehend the conference discourse.

Table 1. Tutors' and Student Writers' Types, Tokens, and Type/Token Ratios.

	Tutors (%)	Student Writers (%)	Tutors and Student Writers (%)
Type	2,922	1,894	3,426
Token	52,589 (70)	22,512 (30)	75,101 (100)
Type/Token Ratio	.055	.084	.045

Frequently used words. The words that occur most frequently in a corpus will not, most likely, constitute the key words that differentiate it from other registers. However, because they can, to some extent, shed some light on what a corpus is about, we discuss them briefly here. We compared the most frequently occurring words that appear in tutors' talk, student writers' talk, and the reference corpora, MASC and COCA. Similar to Farr's (2007) findings from interactions

between tutors and student teachers, we found that the 10 most frequent words in tutors' talk also appeared in the 20 most frequent words of the reference corpora. Overall, 12 of tutors' 20 most frequent words fell into the reference corpora's 20 most frequent words. Student writers' talk mirrored the reference corpora's 20 most frequent words as well. Of student writers' 20 most frequent words, 13 corresponded to the reference corpora's 20 most frequent words. Of student writer's seven most frequent words that fell outside of the reference corpora's 20 most frequent, two were minimal responses ("ok" and "yeah") and two were hesitation markers ("like" and "um"). These findings suggest that student writers signaled they were following along and attending to what the tutor said. They also suggest that linguistic means for slowing discourse in order to think and respond appropriately characterize student writers' talk. As was the case for the student teachers in Farr's (2007) research, both their minimal responses and their hesitation markers coincide with their role as less-expert writers and help seekers.

Tutors and student writers together used 11 most frequent words that fell outside the reference corpora's 20 most frequent words. (See Table 2. Words not occurring in the 20 most frequent words in the reference corpora are bolded.[4]) These were the most frequent words that contributed some sense of the aboutness of our corpus. But, as we predicted, tutors' and student writers' most frequent words (unlike their key words, discussed below) did little to distinguish their talk from the reference corpora; for the most part, their most frequent words constituted the most frequent words in English—no matter the register.

That said, in contrast to tutors' most frequent words, 2 of student writers' 10 most frequent words ("like" and "ok") fell outside the 20 most frequent words in the reference corpora. As mentioned previously, "ok" can function as a backchannel, indicating that the listener is following along with the speaker. Researchers have discussed "like" as a hesitation marker, and indeed it makes sense to classify it as one; however, recent research reveals how "like" indicates a speaker's acknowledgement of imprecision in what he or she is saying. Gisle Anderson (2000) puts it this way: "'Like' signals a slight discrepancy between the speaker's utterance and what the speaker has in mind, or between a state of affairs and the speaker's description of it" (pp. 44–45). In addition, "like" has become more common as it has grown to have a quotative use in the last 20 to 25 years (for example, Ferrara & Bell, 1995; Dailey-O'Cain,

4 We did not include tutors' use of "are" as it is a form of "be" like "is" in the list of 20 most frequent words in the reference corpora.

2000). As we discuss in more detail later, these two words differentiate student writers' talk in writing center conferences from other talk.

Table 2. Most Frequent Words in the Talk of Tutors and Student Writers and in the Reference Corpora.

Tutors			Student Writers			Reference Corpora		
	Freq	Type		Freq	Type		Freq	Type
1	2,822	you	1	1,329	i	1	20,637	the
2	1,783	that	2	713	**like**	2	12,029	to
3	1,666	the	3	645	that	3	11,250	and
4	1,506	to	4	613	it	4	10,586	of
5	1,420	is	5	612	is	5	9,882	a
6	1,155	and	6	611	the	6	9,559	that
7	1,093	i	7	568	and	7	9,533	i
8	1,053	it	8	566	to	8	7,629	you
9	1,044	of	9	430	**ok**	9	7,404	s
10	925	a	10	393	not	10	7,139	in
11	873	**ok**	11	390	**yeah**	11	6,966	it
12	802	**so**	12	341	a	12	5,741	is
13	713	are	13	319	of	13	4,521	we
14	668	**what**	14	313	**um**	14	3,593	for
15	607	this	15	298	**just**	15	3,520	this
16	603	**do**	16	290	you	16	3,423	was
17	578	not	17	260	**do**	17	3,261	they
18	552	**like**	18	251	**so**	18	3,154	he
19	545	**know**	19	250	this	19	3,129	t*
20	543	**have**	20	210	in	20	3,019	on

*t in the reference corpora stems from the contraction and means "not."

Key words. As Andrew Kehoe & Matt Gee (2011) said, one of the best ways to understand the aboutness of writing center talk (or any specialized corpus) in order to discover what makes it unique is to compare its key words to a reference corpus (or corpora). As noted before, a key word is more than a frequently occurring word. It is a word that occurs with statistical significance more or less frequently in comparison to other spoken or written texts (gauged through a reference corpus or corpora). For example, a medical term such as "bunion" might be a key word in the corpus of podiatrist office discourse, but "bunion" would be less likely to materialize in a corpus of everyday conversation. We used AntConc to cross-tabulate types and tokens in the writing center corpus and in the reference corpora, resulting in a chi-squared statistic that expressed the extent to which a particular word occurs unusually frequently in our corpus as compared to the reference corpora. (See Table 3 for the key words in tutors' and student writers' talk. Writing-related key words are bolded.) The larger the chi-squared statistic, the more key the word to writing center talk and its aboutness.

Table 3. The 15 Most Key Words in the Talk of Tutors and Student Writers.

Tutors				Student Writers			
	Freq	Keyness	Type		Freq	Keyness	Type
1	873	2922.026	ok	1	713	1899.275	like
2	2822	2384.975	you	2	430	1817.206	ok
3	471	1326.271	um	3	313	1214.331	um
4	802	752.100	so	4	172	1061.231	uhhuh
5	1420	605.733	is	5	1329	1019.971	i
6	309	598.190	kind	6	390	957.059	yeah
7	552	551.011	like	7	142	463.163	am
8	**104**	**466.444**	**thesis**	8	393	401.743	not
9	**118**	**459.841**	**sentence**	9	298	343.161	just
10	**131**	**429.422**	**paper**	10	612	286.325	is
11	713	410.619	are	**11**	**62**	**253.281**	**paper**
12	81	371.923	uhhuh	12	113	222.529	kind

Tutors				Student Writers			
13	**80**	**367.331**	**comma**	13	613	176.197	it
14	668	358.113	what	14	92	175.798	uh
15	1783	347.990	that	15	166	163.703	because

As Table 3 shows, our key-word results show that minimal response words, particularly "ok" and "uhhuh," and hesitation words, particularly "like" and "um," characterize student writers' talk. These words, especially "ok" and "um," also mark tutors' talk. These words differentiated our writing center corpus from the reference corpora.

The results of our analysis of key-words (as well as the results of our analysis of frequently occurring words) correspond to findings from previous writing center research that characterizes writing center discourse as asymmetrical collaboration, where tutors encourage active participation from student writers and interactivity is vital. Susan Blau, John Hall, & Tracy Strauss (1998), for example, included backchannels in their discussion of echoing (p. 22). Terese Thonus (2002) referred to backchannels, along with laughter and overlaps, as "interactional features" (p. 121). She found that low rates of backchanneling, among other interactional features, signaled low involvement in conferences.

Besides these minimal responses and hesitation markers, in tutors' talk, the reference word "you" was second most key. Clearly, tutors focused their attention on student writers (the "you" in their discourse). Student writers' talk focused on themselves; the pronoun "I" was the fifth-most key word in their talk. The keyness of these pronouns indicates the institutional, service-oriented nature of writing center talk, coinciding with Farr's (2007) key word findings from her comparison of the POTTI corpus with the Limerick Corpus of Irish English. Unlike her comparisons of frequently occurring words, Farr did not analyze tutors' and student teachers' talk separately in determining key words, so she found two key reference words: the pronoun "I" was third-most key, and the pronoun "you" was fourth-most key (p. 248). The nature of a dyadic tutoring interaction is such that the participants will be inwardly focused, concentrating on the student writer.

In using key words to understand the aboutness of writing center conferences, we saw that tutors' talk in particular contained a variety of key writing-related words. "Thesis," "sentence," "paper," and "comma" (eighth-, ninth-, tenth-, and thirteenth-most key) were more key to tutors' talk than to student writers' talk. Such metadiscourse words signal, in Etienne Wenger's (1999) terms, a community of practice in that

they compose a specialized vocabulary that community members speak. Farr (2007) also found metadiscourse words that characterized the talk in her conferences, words related to writing and teaching, for example, "lesson," "class," and "words" (p. 249), and, as mentioned previously, DaSilva & Rennick (2010) found that by the third PBL session veterinary students had increased their use of appropriate medical vocabulary, one indication of learning. In our study, metadiscourse related to writing and the writing process, and tutors and student writers differed in their use of it. Tutors' talk, as discussed before, focused on thesis statements. It also focused on discrete sentences, on the paper as a whole, on the paper as an assignment, and on student writers' use of commas. Student writers' talk, on the other hand, was less characterized by metadiscourse overall. When they did use metadiscourse, they focused on their papers as class assignments. The most key metadiscourse word in their talk was "paper" (eleventh-most key).

To delve deeper into the aboutness of writing center conferences through tutors' and student writers' key words, we examined the context of their key words; the key word became a key word in context. For tutors, a critical aim of many conferences is making sure that student writers have clearly articulated a main point. (See Figure 1 for a screenshot of "thesis" in context.) Tutors used the word "thesis" to help student writers' craft a thesis statement that they could support: "Like when you create your thesis, there might be a phrase in there, you know, 'to a certain extent increases' and then you cover yourself" (10T). They also used it when helping student writers create more comprehensible thesis statements: "What if we change one of the awkward verbs in your thesis, you know, 'should be informed'?" (11T). They also used "thesis" to discuss the relationship of supporting paragraphs to the main point: "Relate that particular example that you're providing back to the thesis" (17T). The concordance tool helped determine the ways that tutors used the word.

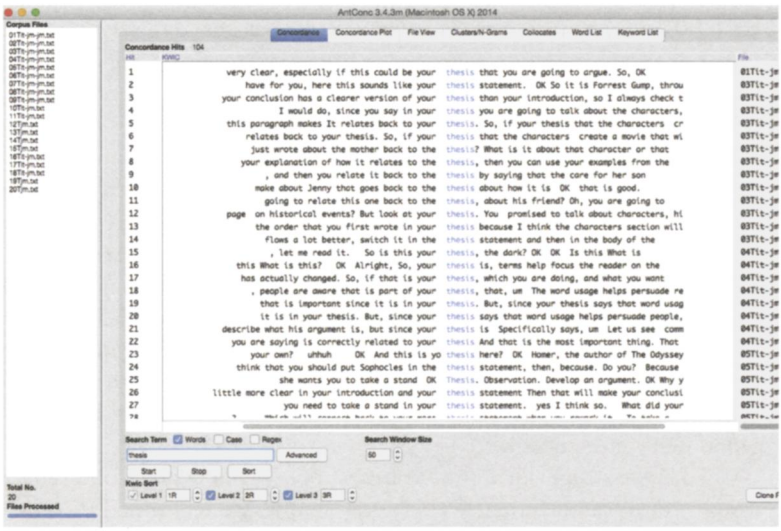

Figure 1. Results for tutors' use of "thesis" in the concordance tool of AntConc.

As noted before, the word "paper" was key in both tutors' and student writers' discourse. For tutors, "paper" was tenth-most key; for student writers, "paper" was eleventh-most key. We found that student writers used the term to ascribe names (presumably learned in class) to their assignments:

20S: It's an analytical paper.

4S: We're writing our first argumentative paper.

17S: I have an observational paper.

They also used the word as a synonym for "assignment":

12S: That's all she gave us about the paper.

16S: You know I told you I got a D on my first paper?

Using the concordance plot tool in AntConc, we were able to visualize student writers' use of "paper" across their conferences. Visualizations of "paper" and other writing-related terms helped us understand when in the progression of the conference student writers and tutors tended to use key writing-related words. Figure 2 shows a screenshot of 8 of the 16 conferences in which student writers used "paper." The 16 plots show a tendency for student writers to use the word toward the beginning of their conferences—the time that they explained their assignments to their tutors.

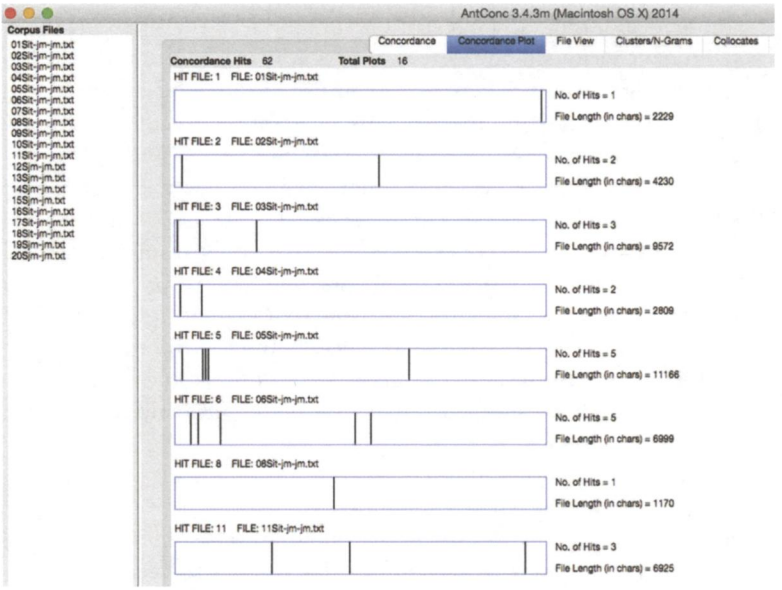

Figure 2. Concordance plot of "paper" in student writers' talk.

For tutors, "paper" was a key word as well. Tutors used it to mean "assignment," like student writers: "And the other paper was thirty percent?" (13T). Not surprisingly, they also used it to mean a sheet of paper: "Why don't we start on a separate piece of paper?" (16T). More interestingly, they used it to refer to the student writer's project holistically: "Tell me what you're thinking as far as main arguments in the paper" (10T). Clearly, discussion of "paper" in all senses of the word differentiates writing center talk from everyday conversation.

The conjunction "so" was fourth-most key in tutors' talk and sixteenth-most key in student writers' talk. The importance of this word in their discourse, particularly in tutors' talk, suggests the ways that tutors and student writers—particularly tutors—establish and develop conclusions from them. For example, tutors used "so" to conclude the discussion of student writers' projects with an evaluation on a global level: "So you did a good job on this" (2T). They also used it to conclude with an evaluation at a local level: "So we've got a nice conjunctive" (15T). The word "so" was key to student writers' talk to a lesser extent, but as the sixteenth-most key word for them, it was important as well. They used "so" to solicit global-level evaluations: "So do you think I

should be okay?" (19S). They also used it to conclude their reasoning at a more local level: "So I cut that out and put it in the thesis" (11S).

Student writers used "not," a negation word, so often that it became their eighth-most key word. Their use of the word differentiated their talk from tutors' and from the discourse of the reference corpora. The extent to which negation indicates the aboutness of student writers' talk relates somewhat to their use of "not" in the frequent lexical bundle "I don't know," discussed as a 4-gram below. Of the 393 occurrences of "not," 57 occurred in the collocation "I don't know." Using AntConc's collocate tool, we found that "I" and "do" and "know" were the three most frequently collocating terms for "not" in student writers' talk when analyzed in a three-word window on the left and the right of "not."

Beyond student writers' use of "not" in "I don't know," they also used negation to delineate their instructors' preferences and restrictions:

12S: She doesn't want contrast.
8S: She said don't summarize.
6S: He didn't say I could not do that.

Such examples often came during the opening stage, the stage during which tutors elicited "what student writers wanted to accomplish (or, at least, thought they wanted to accomplish)" and "determined what the final product should be and where student writers were in the composing process" (Mackiewicz & Thompson, 2015, p. 64). In these situations, the student writers used negation as they tried to articulate the boundaries on their assignments.

Student writers also used negation to list steps in the writing process that they had not yet accomplished:

10S: I couldn't find too much research on that.
4S: I haven't done my works cited yet.
11S: I haven't added them in yet.

The keyness of student writers' use of "not" in examples such as these and elsewhere during their conferences gives a sense that negativity—at least linguistic negativity—pervades their discourse.

Student writers' frequent use of the word "just" differentiated their talk from that of tutors and from the reference corpora. The word was ninth-most key in their talk. Student writers' frequent use of "just" makes sense, given the word's possible meanings. They used the word in several ways:

- To mean "recently": "And because I just said 'because,' do I want that?" (19S)
- To mean "simply": "I'll just have to come back some more." (6S)

- To mean "only": "So I'm just going to get rid of this one." (17S)
- To mean "exactly": "My parents were just like 'You're changing schools.'" (13S)

That student writers would need to point to the recent past—recent revisions to their papers and to content in those papers to which they recently referred—makes sense. So too do their uses of "just" to get at the essence of something ("simply"), to point out one thing (such as a topic) out of others ("only"), and to state something precisely ("exactly"). All of these meanings of "just" set limits on the utterance that encloses them; attempts to single out and articulate their meaning characterizes student writers' discourse.

In this section, we've examined tutors' and student writers' frequently occurring words, type/token ratios, and key words in order to illustrate how corpus-driven analysis can help reveal the aboutness of writing center talk. In the next section, we build on this analysis by examining the words that occurred in proximity to tutors' and student writers' writing-related key words: their key word collocations.

Collocations of writing-related key words. Another way to understand the aboutness of writing center talk is to analyze the words that occur in the environment of the previously determined key words. We analyzed the words that collocated, or co-occurred, with tutors' and student writers' writing-related key words. The collocation tool supplies a mutual information (MI) score for a word within a certain window around another word. Based on the number of times the two words appear together versus separately, MI score measures the strength of association between the two words. To determine the words that strongly collocated with key writing-related words, we set the minimum frequency of co-occurrences to three (to weed out words that occurred infrequently but happened to occur in the environment of the key word that we were analyzing), and we set the window for collocating words at three words to the right and three words to the left. In this section, we discuss the collocating terms for the key word "paper"—a key word that tutors and student writers had in common.

Using these search limits and looking first at student writers' talk, we found that the writing-related key word "paper" collocated with the verb "read" most strongly and with the verb "write" third-most strongly. (Figure 3 shows a screenshot of AntConc's output for collocations of "paper" in student writers' talk.) Both "read" and "write" tended to collocate with "paper" in the opening stage of conferences. For example, "paper" and "read" collocated in 15S's talk when the student

writer conveyed what she wanted her tutor to do: "I just basically need you to read through my paper." Similarly, student writers used "paper" and "write" together to explain the assignment, as in this case from 3S: "All we have to do is- Just have to write a paper about a movie and evaluation." In addition, their use of "write" with "paper," as in this comment from 17S, sheds light on the challenge of writing: "They supposedly tell you how to write a paper in high school, but they don't."

"Paper" also collocated with adjectives that student writers used to describe their papers—"last," "first," and "whole"—and with collocating words indicating the way student writers conceptualized their current work in relation to other writing assignments and their ownership of that work.

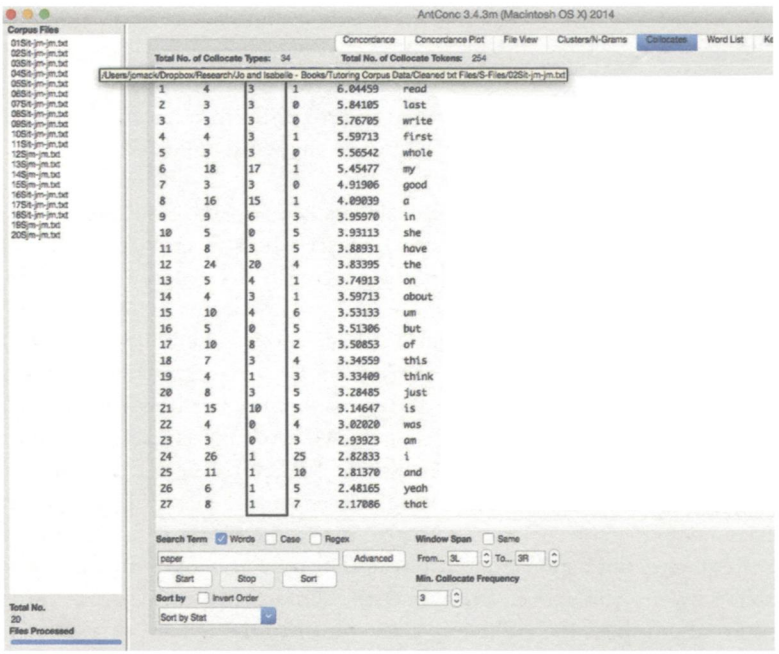

Figure 3. Collocating Words (via MI Score) of Student Writers' Use of "Paper."

Comparing these collocations and their MI scores to those of "paper" in tutors' talk reveals an important limitation of MI score. The highest MI scores for collocating words of "paper" in tutors' talk were "gives," "piece," and "eighty." (See Figure 4 for a screenshot of collocating words of "paper" in tutors' talk.)

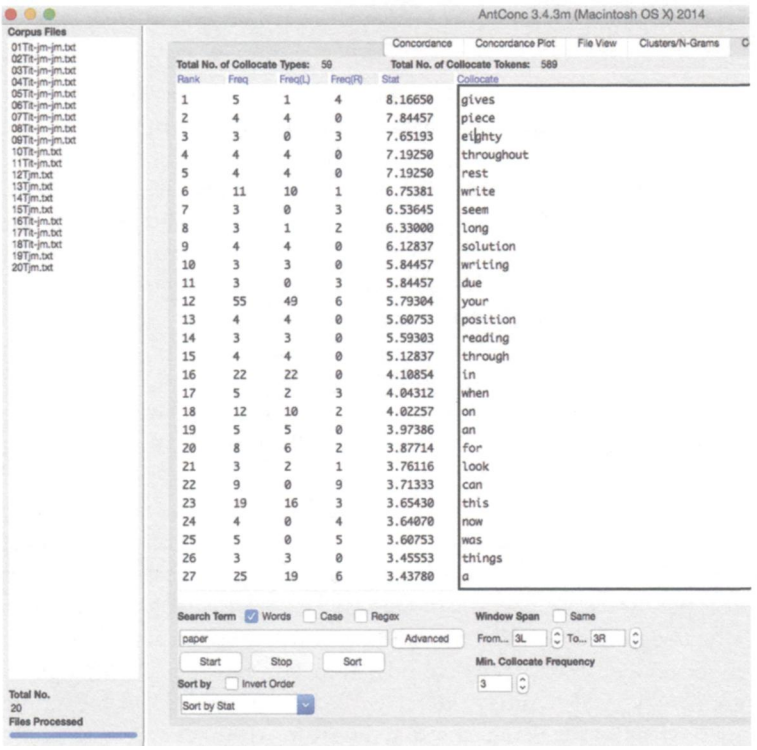

Figure 4. Collocating Words (via MI Score) of Tutors' Use of "Paper."

Two of these collocating words ("gives" and "eighty") occurred in just one conference (14T–14S). These results for the association between "gives" and "eighty" and "paper" reveal a problematic characteristic of MI score: It will give a deceptively high score to any pair of words "for which the frequency of co-occurrence is a high proportion of the overall frequency of either of the pair" (Collins Wordbank). Thus, words such as "give" and "eighty"—words that occurred infrequently overall—generated a high MI score. One solution to this problem with low-frequency collocating words is to calculate a *t*-score—a score that AntConc supplies as well. *T*-score measures the likelihood that two words are associated, as opposed to the strength of association between two words. *T*-score "is good for more grammatically conditioned pairs, like 'depend on'" (Collins Wordbank). Determining *t*-score, we found the words "your," "the," and "is" ranked the highest as collocates of

"paper" in tutors' talk. *T*-score, then, can serve as a useful supplement to MI score under certain conditions, but it is less useful for understanding aboutness because it can produce a high score for words that occur together in any context as opposed to words that occur together in a specific type of discourse, such as our specialized corpus of writing center conferences.

Our discussion of the collocating words for the writing-related key word "paper" illustrates how writing center researchers can use a corpus-driven approach to identify associations between important words that might produce insights into tutors' and student writers' talk through subsequent qualitative analysis.

Lexical bundles (n-grams). To better understand the identifying features, the aboutness, of our study corpus, we analyzed 4-grams, bundles of words such as "I don't know." Citing Douglas Biber, Stig Johansson, Geoffrey Leech, Susan Conrad, & Edward Finegan (1999), Lukasz Grabowski (2015) defines a lexical bundle as three or more words "that occur frequently in natural discourse and constitute lexical building blocks" (p. 25). Grabowski's definition suggests a view that lexical bundles are sequences of language that speakers produce holistically as opposed to word for word, linearly. Researchers differ in a number of occurrences they consider sufficient in order to call a given n-gram a "lexical bundle." Here, we bypassed this debate by simply using the terms interchangeably and discussing the n-grams, specifically the 4-grams, that occurred most frequently in our corpus.

We analyzed 4-grams (as opposed to 3- or 5-grams) because, as Eniko Csomay (2013) points out in her corpus analysis of classroom discourse, "three-word bundles are often too prevalent and have proven difficult to interpret" and "there are too few five-word bundles" (p. 371). Below, we discuss the three most frequent 4-grams in tutors' talk and the three most frequent 4-grams in student writers' talk. Figure 5 shows a screenshot of AntConc's output for tutors' 4-grams.

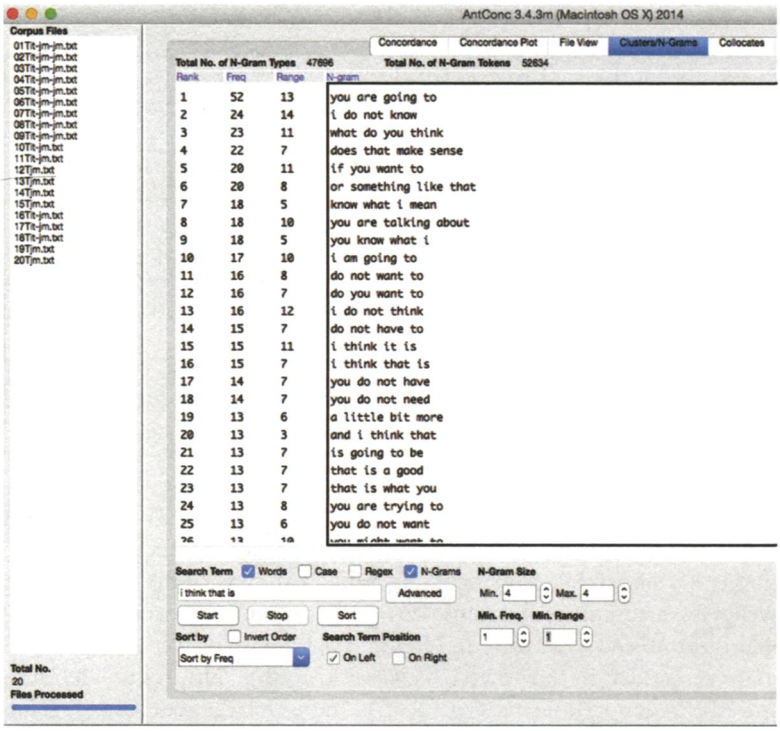

Figure 5. 4-grams in tutors' talk.

As Table 4 shows, both tutors and student writers used one 4-gram far more frequently than the other 4-grams that they used: tutors used "you're going to" 52 times across 13 conferences, and student writers used "I don't know" 57 times across 15 conferences.

Table 4. Frequency and Range of Tutors and Student Writers' 4-Grams.

Tutors	Freq	Range	4-gram	Contracted Form
	52	13	you are going to	you're going to
	24	14	I do not know	I don't know
	23	11	what do you think	what do you think
Student Writers	57	15	I do not know	I don't know
	15	8	I am going to	I'm going to
	14	13	that is what I	that's what I

Before discussing the frequency of tutors' and student writers' 4-grams in more depth, it is helpful to analyze the syntactic structure of these frequently occurring sequences. As Table 4 shows, tutors' and student writers' most frequent 4-grams were fairly consistent in structure. They mainly consisted of a pronoun ("I" or "you") plus an auxiliary verb and main verb. That is, they shared a clausal structure. This result—a clause structure—is not, it seems, typical of 4-grams in larger corpora. Talking about lexical bundles, Sylviane Granger (2014) points out that they can take the form of phrases or clauses. In corpora that are less specialized than ours, a frequent 4-gram might be a phrase such as "and things like that." The clausal structure of tutors' and student writers' 4-grams suggests that these bundles served as clausal frames. Tutors or student writers could begin or lead off with "You're going to" or "I'm going to" and then fill in the subsequent slot (for example, T3's instruction, "So then you're going to have to back it up").

In addition, n-grams, including 4-grams, range in function as well. Granger (2014) cites Biber, Conrad, & Viviana Cortes's (2004) delineation of three main types of bundles in terms of function: referential bundles, such as "in the United States"; discourse organizers, such as "with this in mind"; and stance bundles, such as "I don't want to" (p. 59). Although tutors' and student writers' 4-grams for the most part shared clausal structure, their functions varied. Tutors' use of "what do you think" clearly connected prior discourse with discourse to come, but 4-grams such as "I don't know" marked stance (Biber, Conrad, & Cortes, 2004, p. 384). Tutors' and student writers' most frequent

4-grams indicate that they monitored the ongoing discourse, iteratively referenced that talk, and continually indicated their stance toward it.

As noted above, the most frequently occurring 4-gram in tutors' talk was "you're going to." This 4-gram occurred 52 times across 13 conferences. This finding is in line with research that shows "be going to" bundles occur more frequently in spoken discourse than in written discourse (Berglund, 1997, p. 16) and such occurrences are "spreading over time" (Szmrecsanyi, 2003, p. 296). Analyzing "you're going to" in context reveals that tutors used it during instruction, as when 3T told 3S what he needed to add to his paper: "You're going to want a transitional statement here that says- something." Besides occurring during instruction sequences, this 4-gram also occurred when tutors responded as readers to student writers' texts. They used the 4-gram to forecast what they expected to read next given what they had just read, as 12T did: "'Some' kind of makes me feel that you're going to tell me immediately"; and 9T did as well: "So you've already got your thesis going here. You're not just going to talk about the ads, but you're going to prove how they're-." This latter use of "be going to" coincides with Biber, Conrad, & Cortes's (2004) categorization of "be going to" as an intention/prediction stance marker (p. 385). Tutors used this 4-gram to provide instructions for next steps but also to explain how words set up expectations in readers' minds—expectations that writers should try to meet.

The intention/prediction stance marker "be going to" appeared in student writers' talk as well. "I'm going to" was the second-most frequent 4-gram for student writers; however, it occurred just 15 times across 8 conferences. Student writers used this 4-gram to assert their next steps for their papers after the conference:

3S: I'm going to revise it.

4S: I'm going to go home and edit.

13S: I'm going to edit that whole out.

In student writers' talk, "be going to" suggests student writers' concern for the tasks ahead of them that will help them develop their writing projects.

As noted before, for student writers, by far the most frequently occurring 4-gram was "I don't know." This 4-gram occurred 57 times in 15 out of 20 conferences. This negative-stance bundle is extremely common across American and British English; in fact, as Nicole Baumgarten & Juliane House (2010) point out, "'I don't know' is the most frequently occurring negative word bundle in both the Corpus of Contemporary American English and the British National Corpus (p. 1186). Its ubiquity arises from its myriad meanings. Beyond its

primary meaning—insufficient knowledge—"I don't know" has other, pragmatic uses: "avoiding assessment, prefacing disagreement, avoiding explicit disagreement and commitment, minimizing impolite beliefs, and indicating uncertainty" (Baumgarten & House, 2010, p. 1194). In their analysis of writing center conferences, Blau, Hall, & Strauss (1998) discuss "I don't know" as a qualifier, sometimes to mitigate directiveness and other times to hedge and provide time for tutors to think through their responses. Prior research, then, suggests that student writers (and tutors) likely have other motivations besides insufficient knowledge for using "I don't know."

Indeed, a closer look at the ways in which student writers used the 4-gram "I don't know" mitigates the worry arising from the interpretation that student writers repeatedly assert their lack of knowledge. Student writers used "I don't know" pragmatically in what Baumgarten & House (2010) call a "verbal routine" to express "uncertainty and avoidance of full commitment to the upcoming or preceding proposition" (p. 1195). A telling example is this one in which 15S expressed uncertainty in her choice of the word "general" as she justified the way she had worded a sentence:

> 15S: I really don't want to put something so specific like "I got back into horseback riding" or something like that. That's just kind of- Because it takes away from- I just like how it's so- I don't know. It's general in a way that people can relate to it. And as soon as I attach something really, really specific to it, I don't think it does that anymore.

When student writers used this common 4-gram, then, they did not necessarily mean that they did not understand or did not have an answer.

That is not to say, however, that student writers never used "I don't know" when they truly had hit a wall in their understanding. 17S, for example, used "I don't know" in its core sense: "I don't know how to begin this one. This one is the one speaking of, like, the similarities about the fourth time." The excerpt below also exemplifies the core sense of "I don't know":

> 12T: What do you mean?
>
> 12S: I mean. I know- I don't know what I mean to be honest.

In cases like these, student writers used "I don't know" to convey their lack of understanding, but such cases do not tell the entire story of student writers' use of this 4-gram.

Tutors also used the 4-gram "I don't know" with some frequency; it was their second-most frequent 4-gram, occurring 24 times across 14 conferences (fewer than half as many occurrences of their most frequent 4-gram, "you're going to"). Tutors, like student writers, used "I don't

know" to show uncertainty and avoid what Baumgarten & House (2010) call an "unequivocal stance" (p. 1194). They used "I don't know" particularly when they were trying to conjure a word or phrase when making a suggestion about wording, as 11T did: "Maybe, like, this kind of, I don't know, ubiquitous control." They also used it as they tried to come up with an example, as 6T did: "For example, if somebody suggested something related to, I don't know, sports." Also like student writers, they used the 4-gram when they sincerely appeared to lack sufficient knowledge of a topic, as 20T did: "So yeah I think it's easier to believe because- I don't know. I can't remember. I just blanked out." But as was the case with student writers' use of "I don't know," tutors used "I don't know" in other, less obvious ways.

Tutors' third most frequent 4-gram, the question and question opener "what do you think," occurred 23 times across 11 conferences. This 4-gram is one that Biber, Conrad, & Cortes (2004) would categorize as an introduction/focus discourse organizer (p. 386). Indeed, tutors used "what do you think" to ask cognitive scaffolding questions. Such questions tend to push student writers to think about their work. But as we have previously discussed (Mackiewicz & Thompson, 2014), cognitive scaffolding questions differ in the extent to which they invite a substantive response. For example, 5T's "What do you think?" could have received a very brief response from 5S, such as "It's probably not." As it turned out, however, 5S followed up with a fairly substantive response—a common ground question:

5T: And I'm just wondering if "effective" is the right word there. What do you think? Or-
5S: You mean like "affective" versus "effective"?
5T: No. "Effective"- No. That would be a- the- that would be the um- the right choice for "affect" or "effect."
5S: Ok.
5T: But is- (3 seconds)
5S: Um.
5T: I was just wondering if there's a stronger way you can say that.
5S: Yeah. Um.

While in this case "What do you think?" left open the possibility of a rather limited response from 5S, it at least prodded the student writer to consider her word choice and to respond with more than a "yes" or "no" answer. Though she did not have to, 5S fully engaged in the conversation by checking to see whether 5T's question related to a usage rule (the distinction between "affect" and "effect") that she had likely encountered before. 5T and 5S went on to discuss further the possibil-

ity of replacing "effective" with a different word without making the sentence repetitive. 5T's "what do you think?" began the process of exploring this option.

The next example shows how a tutor's "what do you think" question left a slightly wider range of possibility for the student writer's response. 19T's question generated a more substantive response from 19S:

19T: Ok. So what do you think the main point of that paragraph is?
19S: Um. (3 seconds) That police aren't doing their job, I guess.
19T: [clears throat] Police aren't doing their jobs because- What?
19S: Because they're basing it solely on race instead of what the citizen-

With 19T's pumping, 19S articulated the main point of the paragraph and with prompting, elaborated her claim about police officers who engage in racial profiling. After a pause and 19T's suggestion to "go back to that topic sentence and think about how this topic sentence can kind of forecast what's about to come in this paragraph," 19S finishes reworking her claim:

19S: Ok. So [reading] "The police shouldn't [unclear] when trying to find a suspect to a crime because-" Um.
19T: Think about what the problem with [this is
19S: [Because they should- Because they should um- (5 seconds) It's the lazy- It's the easy way out?
19T: Sure.

In this case, 19T's scaffolding question sparked a line of substantive responses from 19S as the two tried to nail down the argument that 19S wanted to make in the paragraph.

The two examples above show that "what do you think" (along with other cognitive scaffolding questions) help tutors push student writers to respond to varying degrees, but as these examples indicate, although the student writers' responses could be fairly brief, they could not be a simple "yes" or "no."

The frequency with which tutors used the question opener "what" and the question "what do you think" corresponds to findings from previous research that reveals the variety of functions that questions serve in writing center conferences. As Willa Wolcott (1989) argued, questions are "a central means for engaging students in dialogue" (p. 20). Three studies of questioning in writing center conferences (Blau, Hall, & Strauss, 1998; Corbett, 2011; Mackiewicz & Thompson, 2015) are particularly useful in understanding the three "what" questions

discussed in this section: "What do you think," "So, what do you think the main point of that paragraph is?" and "What?" According to the scheme of tutoring strategies we laid out in earlier research (Mackiewicz & Thompson, 2014), all three "what" questions serve as cognitive scaffolding, distinguished in the three instances by the (context-dependent) constraints imposed on the responder. "What do you think?" is minimally constraining if the tutor has only a general notion of an acceptable response. Tutors are likely to have more specific responses in mind when they ask questions such as "What do you think the main point of that paragraph is?" and "What" in the example above is likely highly constraining, called a prompt (Mackiewicz & Thompson, 2014). Steven J. Corbett (2011) would likely classify these three questions as "clarifying," aimed at helping student writers clarify their thinking, and "open-ended," aimed at allowing a wide range of responses from student writers (p. 65). Blau, Hall, & Strauss (1998) would likely identify the purpose of all three questions as eliciting responses from student writers and by type either as open, inviting discussion, or closed, inviting a correct response (p. 23). Rather than gaining information they need from students, tutors ask these three "what" questions to help student writers more forward in their thinking.

To finish up with the most frequent 4-grams that tutors and student writers used, here we look at student writers' third most frequent 4-gram. "That's what I" occurred 14 times across 13 conferences. One way that student writers used this 4-gram was to ratify their tutors' questions, as in this example from the 18T–18S conference:

18T: So, in your class did you talk about the human condition? Like, have you-

18S: I'm not quite sure. That's what I was kind of wondering- What that she meant by that.

And this example from the 1T–1S conference:

1T: So um, which one were you more interested in?

1S: That's what I'm having a hard decision with. Like, picking.

But they also used it to convey their understanding of or intention in previous discourse:

20S: And once you get to the resolution, it's like in order to- Creation, you know, put humans on earth. It's kind of- That's what I got from it.

5S: I mean I was going for- The gods want to think that they're fighting all the time like a person. That's what I want it to say.

Biber & Federica Barbieri (2007) and Biber, Conrad, & Cortes (2004) label similar lexical bundles as referential and identification/focus

because they direct attention toward other language in the interaction. As Biber & Barbieri point out, such lexical bundles can be forward looking, focusing on the noun phrase that follows. In contrast, "that's what I" pointed backward to what tutors had said, to what they read in a text, or to what they had written in their papers or stated previously rather than looking forward to forthcoming speech. This 4-gram appeared frequently, then, because student writers needed to determine what they meant in the first place and often to rearticulate what they meant.

We conclude in this section on tutors' and student writers' 4-grams that even a fairly brief analysis of frequently occurring 4-grams provides insight into tutors' and student writers' talk. For example, their frequent use of the 4-gram "I don't know" sheds light on their stance, specifically their certainty (or lack of it) about what they were saying, in addition to their admissions of a lack of understanding. Like the frequently occurring words and key words, the 4-grams derived from our corpus-driven study point to other empirical research about writing center talk. As would be predictable in expert–novice asymmetrical collaborations, tutors' 4-grams reflect their role of expert advice giver ("you're going to"), but they also reflect the writing center lore to be as nondirective as possible ("what do you think"). Student writers' use of the 4-grams "that's what I" and "I'm going to" indicate how student writers' talk refers backward and forward in time. With "that's what I," they acknowledged expert (tutor) questions and conveyed their understanding and intentions. With "I'm going to," student writers looked ahead to tasks that would complete their writing process. Much discussion about these seemingly contrasting roles has been published (Thonus 1999a, 2002; Corbett, 2011; Mackiewicz & Thompson, 2015).

Conclusion

This article demonstrates a combination of quantitative corpus linguistic analysis and qualitative discourse analysis can provide a true mixed-methods approach to understanding writing center conferencing, particularly the ways that tutor and student-writer talk differs from other registers. In discussing the importance of qualitative data collection and analysis in educational research, Frederick Erickson (1986) argued for the importance of what he called "interpretative" research as a means for addressing the "*invisibility of everyday life*" (p. 121; italics in the original). For many years, writing center researchers have explored Erickson's admonition about the importance and power of discourse analysis to examine writing center talk through detailed descriptions of individual conferences and comparisons of two or more conferences (for

example, Davis, Haywood, Hunter, & Wallace, 1988; Wolcott, 1989; Kiedaisch & Dinitz, 1993; Blau, Hall, & Strauss, 1998; Brown, 2010). However, corpus-driven research can also make the invisible visible. It can achieve this goal by uncovering key words, sequences of words, and other linguistic features so commonly used in writing center dialogue that researchers may not single them out for analysis and by using those features to differentiate writing center language from other spoken language. With a representative sample of the writing center register, corpus linguistic analysis can yield findings generalizable beyond the local, something qualitative research alone cannot do. When used together, corpus analysis and discourse analysis can provide a thorough look at naturally occurring language.

Our example of corpus-driven analysis provides insights into the aboutness of our 20 writing center conferences by bringing to the forefront what tutors and student writers talked about most often. Overall, our findings support and add to current research about what is discussed in writing center conferences and how. As mentioned previously, they support the view that writing center conferences are asymmetrical, with tutors assuming the more expert role and student writers assuming the less expert role. However, our analysis tempers this portrayal of lop-sided power by revealing the tutors' concern to develop rapport with the student writers and to support them in achieving their goals.

Specifically, we report five groups of findings from our corpus-driven study. Our first group shows that tutors' talk manifested limited the lexical variety and focused in on the subject at hand, the writing issues that bring students to the writing center. Hence, the low type/token ratio of the tutors' talk may be interpreted as a reflection of their consistent concern to serve students rather than impress them. Second, our comparison of the most frequently occurring words in our writing center corpus to those in two reference corpora indicates only a few differences in word usage. Hence, as early writing center practitioners noted (for example, Bruffee, 1984) writing center talk shares commonalities with other spoken language, including casual conversation.

The third group of findings, the key words in our writing center corpus, is probably the most revealing. The key words identified indicate the collaborative nature of our conferences, through the minimal responses, particularly backchannels, that the tutors and student writers used to support floor holding for each other. Further, they clearly identified that the student writers were the focus of the conferences, through the tutors' use of "you" as a key word and the student writers' use of

"I." Key words, particularly those associated with tutors, also include writing-related terms.

The last two groups of findings from the corpus-driven analysis identify aboutness through frequently occurring sequences of words in our writing center corpus. Examining sequences of words occurring with the writing-related term "paper," we found what we might expect: that in the opening stage of conferences, student writers often sequenced "paper" with the term "read." Many more writing-related terms can be considered using MI score. Our analysis of 4-grams shows that "I don't know" can do more than convey a lack of relevant knowledge. However, student writers' far more frequent use of the 4-gram indicates an asymmetry in writing center conferences, an asymmetry also indicated in other commonly occurring lexical bundles.

Creating A Repository for Writing Center Data

The approach that we are now taking to our writing center research—an approach that includes corpus-driven analysis—has made us excited about the research that is possible if we writing center researchers better leverage our data—our transcripts and, potentially (and even better), our audio and video recordings as well. What we have in mind is a repository for data—a managed and secure space for a corpus of writing center data—something like the ones disciplines such as chemistry, physics, and life sciences have built to share and reuse data.

Before discussing issues surrounding such a project, we think it might be important to point out in the first place the benefits of sharing data. Christine L. Borgman (2012) lists four "rationales" for sharing: 1. to reproduce or to verify research; 2. to make the results of publicly funded research available to the public; 3. to enable others to ask new questions of extent data; 4. to advance the state of research and innovation (p. 1072). We agree with each of Borgman's reasons for moving toward a culture of sharing in the humanities and social sciences. In particular, we think her third reason is especially pertinent to writing center researchers: By sharing data, we make it possible to ask and answer new research questions.

As Rick Gilmore, Lisa Steiger, & Dylan Simon (2015) point out as they argue the case for sharing research videos, data sharing in the hard sciences has "demonstrated benefits for scientific transparency and accelerated discovery" (para. 4). In fact, PLoS ONE's editorial policies states that "Authors must follow standards for data deposition in publicly available resources including those created for gene sequences, microarray expression, structural studies, and similar kinds of data" (PLoS

ONE). Although researchers outside the hard sciences are beginning to discuss the benefits of and pathways to data sharing (for example, Karen E. Adolph, Rick O. Gilmore, Clinton Freeman, Penelope Sanderson, & David Millman, 2012), the practice is not at all common.

In addition to the challenge of planning and building a repository for writing center data—and perhaps other data related to communication pedagogy—is the challenge of making the case for reusing data to our universities' Institutional Review Boards (IRBs). The UK's Economic and Social Research Council (ESRC) already requires researchers to consider "the long-term use, including the potential for data linkage and preservation of data, when obtaining consent" (p. 25). For the ESRC, reuse is the goal when "non-sensitive data or data where there is minimal risk of disclosure of the identity of individuals" (p. 25). Such an attitude toward reuse of data would greatly facilitate and enhance writing center and other communication research. But as Samantha Guss (2009) found in her case study of IRB applications at New York University, "Language used on IRB applications (as well as language that is *not* used), and the underlying ethics of the issue seem to be barriers to data archiving" (italics in original). Guss also found that "Neither [researchers nor IRB members are] clear on the 'rules' or if there are any." If we want to share the data that we collect, we need to make our case early on as we develop our studies, including our consent procedures and forms, and our IRB applications.

Easy and secure data sharing would allow researchers to build and access corpora of writing center talk (as well as other communication generated in writing centers) and thus facilitate a broad array of studies. Such a resource would make the mixed-method approach that we advocate in this article even more useful and powerful. Even as it stands, though, the approach we've described—combining corpus analysis with discourse analysis—helps to unfetter the research questions that we can ask about the talk that goes on in writing centers.

References

Adolph, K. E., Gilmore, R. O., Freeman, C., Sanderson P., & Millman, D. (2012). Toward open behavioral science. *Psychological Inquiry: An International Journal for the Advancement of Psychological Theory, 22*, 244–247.

Anderson, G. (2000). The role of pragmatic marker like in utterance interpretation. In G. Anderson & T. Fretheim (Eds.), *Pragmatic markers and propositional attitude* (pp. 17–38). Amsterdam/ Philadelphia: John Benjamins.

Anthony, L. (2014). AntConc (Version 3.4.3) [Computer Software]. Tokyo, Japan: Waseda University. Available from http://www.laurenceanthony.net/

Baumgarten, N., & House, J. (2010). 'I think' and 'I don't know' in English as Lingua Franca and native English discourse. *Journal of Pragmatics, 42,* 1184–1200.

Berglund, Y. (1997). Future in present-day English: Corpus-based evidence on the rivalry of expressions. *ICAME Journal, 21*, 7–20.

Biber, D. (2006). Stance in spoken and written university registers. *Journal of English for Academic Purposes, 5*, 97–116.

Biber, D., & Conrad, S. (2009). *Register, genre, and style.* Cambridge: Cambridge University Press.

Biber, D., Conrad, S., & Cortes, V. (2004). 'If you look at...': Lexical bundles in university teaching and textbooks. *Applied Linguistics, 25*, 371–405.

Biber, D., Conrad, S., & Reppen, R. (1998). *Corpus linguistics: Investigating language structure and use.* Cambridge: Cambridge University Press.

Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *Longman grammar of spoken and written English.* Harlow: Longman.

Blau, S., Hall, J., & Strauss, T. (1998). Exploring the tutor/client conversation: A linguistic analysis. *Writing Center Journal, 26*(1), 19–48.

Borgman, C. L. (2012). The conundrum of sharing research data. *Journal of the American Society for Information Science and Technology, 63*, 1059–1078.

Brown, R. (2010). Representing audiences in writing center consultation: A discourse analysis. *Writing Center Journal, 30*(2), 72–99.

Bruffee, K. (1984). Peer tutoring and the 'conversation of mankind.' In G. A. Olson (Ed.), *Writing centers: Theory and administration* (pp. 3–15). Urbana: IL: NCTE.

Butler, C. (2004). Corpus studies and functional linguistic theories. *Functions of Language, 11*, 147–186.

Collins Wordbanks. (n.d.). A guide to statistics: T-score and mutual information. *Collins Wordbanks Online.* Retrived from http://wordbanks.harpercollins.co.uk/Docs/Help/statistics.html

Corbett, S. J. (2011). Using case study multi-methods to investigate close(r) collaboration: Course-based tutoring and the directive/nondirective continuum. *Writing Center Journal, 31*(1), 55–81.

Csomay, E. (2013). Lexical bundles in discourse structure: A corpus-based study of classroom discourse. *Applied Linguistics, 34*, 369–388.

Dailey-O'Cain, J. (2000). The sociolinguistic distribution of and attitudes toward focuser "like" and quotative "like." *Journal of Sociolinguistics, 4*, 60–80.

DaSilva, A. L., & Dennick, R. (2010). Corpus analysis of problem-based learning transcripts: An exploratory study. *Medical Education, 44*, 280–288.

Davis, K., Hayward, N., Hunter, K., & Wallace, D. L. (1988). The function of talk in the writing conference: A study of tutorial conversation. *Writing Center Journal, 9*(1), 44–51.

Economic and Social Research Council (ESRC). ESRC framework for research ethics. Retrieved from http://www.esrc.ac.uk/_images/framework-for-research-ethics_tcm8-33470.pdf

Erickson, F. (1986). Qualitative methods in research on teaching. In M. C. Whittrock (Ed.), *Handbook of research on teaching,* 3rd ed. (pp. 119–161). New York, NY: Macmillan.

Farr, F. (2003). Engaged listenership in spoken academic discourse: The case of student–tutor meetings. *Journal of English for Academic Purposes, 2,* 67–85.

Farr, F. (2007). Spoken language analysis as an aid to reflective practice in language teacher education: Using a specialized corpus to establish a generic fingerprint. In M. C. Campoy & M. J. Luzón (Eds.), *Spoken corpora in applied linguistics,* Vol. 51 (pp. 235–258). Peter Lang: Frankfort.

Ferrara, K., & Bell, B. (1995). Sociolinguistic variation and discourse function of constructed dialogue introducers: The case of be + like. *American Speech, 70,* 265–290.

Gilmore, R. O., Steiger, L. M., & Simon, D. A. (2015, June). Databrary: A digital data library for sharing research video. Paper presented at the Joint Conference on Digital Libraries 2015, Knoxville, TN. Retrieved from https://github.com/databrary/presentations/blob/master/jcdl-15/ms-word/jcdl-15-paper-2015-01-14-001-rog.md

Gozdz-Roszkowski, S. (2011). *Patterns of linguistic variation in American legal English.* Frankfort: Peter Lang.

Grabowski, Ł. (2015). Keywords and lexical bundles within English pharmaceutical discourse: A corpus-driven description. *English for Specific Purposes, 38,* 23–33.

Granger, S. (2014). A lexical bundle approach to comparing languages: Stems in English and French. *Languages in Contrast, 14,* 58–72.

Guss, S. (2009, Dec.). *Social science data reuse and institutional review boards: A pilot study of stakeholders' opinions.* Poster presented at the 5th International Digital Curation Conference, London, UK. Retrieved from http://www.dcc.ac.uk/sites/default/files/documents/events/dcc-2009/programme/posters/poster019guss.pdf

Hardie, A., & McEnery, T. (2010). On two traditions in corpus linguistics, and what they have in common. *International Journal of Corpus Linguistics, 15,* 384–394.

Hunston, S. (2002). *Corpora in applied linguistics.* Cambridge: Cambridge University Press.

Kehoe, A., & Gee, M. (2011). Social tagging: A new perspective on textual "aboutness." *Studies in Variation, Contacts and Change in English, 6.* Retrieved from http://www.helsinki.fi/varieng/series/volumes/06/kehoe_gee

Kiedaisch, J., & Dinitz, S. (1993). Look back and say "so what": The imitations of the generalist tutor. *Writing Center Journal, 14*(1), 63–74.

Mackiewicz, J., & Thompson, I. (2014). *Talk about writing: The tutoring strategies of experienced writing center tutors.* New York, NY: Routledge.

McEnery, T., & Wilson, A. (2001). *Corpus linguistics,* 2nd ed. Edinburgh: Edinburgh University Press.

Pemberton, M. (2010). Introduction to "The function of talk in the writing center conference: A study of tutorial conversation." *Writing Center Journal 30*(1), 23–26.

Phillips, M. (1989). *Lexical structure of text. Discourse analysis monograph,* no. 12. Birmingham: English Language Research. University of Birmingham.

PLoS ONE Publication Criteria. PLoS One. (n.d.). Retrieved from: http://www.plosone.org/static/publication#data report

Reinhardt, J. (2010). Directives in office hours consultations: A corpus-informed investigation of learner and expert usage. *English for Specific Purposes, 29,* 94–107.

Schmitt, N. (2000). *Vocabulary in language teaching.* Cambridge: Cambridge University Press.

Severino, C., & Deifell, E. (2011). Empowering L2 tutoring: A case study of a second language writer's vocabulary learning. *Writing Center Journal, 31*(1), 25–54.

Stuart, K. (2006). Towards an analysis of academic weblogs. *Revista Alicantina de Estudios Ingleses, 19,* 387–404.

Szmrecsanyi, B. (2003). "Be going to" versus "will/shall": Does syntax matter? *Journal of English Linguistics, 31,* 295–323.

Thompson, I., & Mackiewicz, J. (2014). Questioning in writing center conferences. *Writing Center Journal, 33*(2), 37–70.

Thonus, T. (1999a). Dominance in academic writing tutorials: Gender, language proficiency, and the offering of suggestions. *Discourse & Society, 10,* 225–248.

Thonus, T. (1999b). How to communicate politely and be a tutor, too: NS–NNS interaction and writing center practice. *Text, 19,* 253–279.

Thonus, T. (2002). Tutor and student assessments of academic writing tutorials: What is "success"? *Assessing Writing, 8,* 110–134.

Walsh, S., Morton, T., & O'Keeffe, A. (2011). Analysing spoken university interaction: A CL/CA approach. *International Journal of Corpus Linguistics, 16,* 325–344.

Wenger, E. (1999). *Communities of practice: Learning, meaning, and identity.* Cambridge: Cambridge University Press.

Will, S. (2012). Special report: The 100 most common words in spoken and written English. Retrieved from: http://idiomas.astalaweb.com/ingl%E9s/v2/50-palabras-mas-comunes-en-ingles-hablado-y-escrito.pdf

Wolcott, W. (1989). Talking it over: A qualitative study of writing center conferencing. *Writing Center Journal, 9*(2), 15–29.

Wooffitt, R. (2005). *Conversation analysis and discourse analysis: A comparative and critical introduction*. London: SAGE Publications.

About the Authors

Jo Mackiewicz is an associate professor of Rhetoric and Professional Communication at Iowa State University. Routledge will publish her latest book, *The Aboutness of Writing Center Talk: A Corpus-Driven and Discourse Analysis*, in early 2017. She has collaborated with Isabelle Thompson to publish three articles about student writer-tutor dialogue in writing centers and a book, *Talk About Writing: The Tutoring Strategies of Experienced Writing Center Tutors*.

Isabelle Thompson retired from Auburn University in 2012. While she was at Auburn, she was director of the English Center, Auburn's writing center, for 15 years and taught a variety of professional writing courses. Since her retirement, she has collaborated with Jo Mackiewicz to publish three articles about student writer-tutor dialogue in writing centers and a book, *Talk About Writing: The Tutoring Strategies of Experienced Writing Center Tutors*.